Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical Function Shortform (KOOS-PS)

Braaksma, C.; Wolterbeek, N.; Veen, M. R.; Prinsen, C. A.C.; Ostelo, R. W.J.G.

published in
Osteoarthritis and Cartilage
2020

DOI (link to publisher)
10.1016/j.joca.2020.08.004

document version
Publisher's PDF, also known as Version of record

Osteoarthritis and Cartilage 28 (2020) 1525–1538

Osteoarthritis and Cartilage

Review

Systematic review and meta-analysis of measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical Function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical Function Shortform (KOOS-PS)

C. Braaksma † *, N. Wolterbeek †, M.R. Veen †, C.A.C. Prinsen §, R.W.J.G. Ostelo ‡ §

† *Department of Orthopaedic Surgery, Utrecht, the Netherlands*
‡ *Department of Health Sciences and the Amsterdam Public Health Research Institute, VU University, Amsterdam, the Netherlands*
§ *Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, de Boelelaan 1117, Amsterdam, the Netherlands*

ARTICLE INFO

Article history:
Received 20 November 2019
Accepted 5 August 2020

Keywords:
Systematic review
HOOS-PS
KOOS-PS
Measurement properties
Patient reported outcome measures (PROMs)

SUMMARY

Objective: The aim of this systematic review and meta-analysis was to evaluate all evidence on measurement properties of the Hip disability and Osteoarthritis Outcome Score - Physical function Shortform (HOOS-PS) and the Knee Injury and Osteoarthritis Outcome Score - Physical function Shortform (KOOS-PS).

Design: This study was conducted according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews of PROMs. MEDLINE, EMBASE, The Cochrane Library, CINAHL and PsychINFO through February 2019 were searched. Eligible studies evaluated patients with hip or knee complaints and described a measurement property, interpretability, feasibility, or the development of either the HOOS-PS or KOOS-PS.

Results: Twenty-three studies were included. For both questionnaires, the content validity was found inconsistent and the quality evidence was moderate for a sufficient reliability and high for an insufficient construct validity. The HOOS-PS had a high quality evidence of sufficient structural validity and internal consistency (pooled Cronbach's alpha 0.80; $n = 3761$) and low quality evidence of sufficient measurement error and indeterminate responsiveness. Concerning the KOOS-PS, the quality evidence was high for an insufficient responsiveness, moderate for an inconsistent structural validity and internal consistency and low for an inconsistent measurement error.

Conclusions: The inconsistent evidence for content validity implies that scores on the HOOS-PS and KOOS-PS may inadequately reflect physical functioning. Furthermore, there is evidence for insufficient construct validity and responsiveness in patients with knee osteoarthritis receiving conservative treatment. Using the HOOS-PS or KOOS-PS as outcome measurement instruments for comparing outcomes, measuring improvements or benchmarking in patients with hip or knee complaints or undergoing arthroplasty should only be done with great caution.

Review registration: PROSPERO number CRD42017069539

Introduction

In total joint arthroplasty, patient reported outcome measures (PROMs) are widely used to evaluate the effect of treatment on individual patients and for comparative effectiveness research. In addition, the health care industry has become interested in using these instruments as an indicator of quality of care[1].

* Address correspondence and reprint requests to: C. Braaksma, Department of Orthopaedic Surgery, St. Antonius Hospital Utrecht, P.O. Box 2500, Nieuwegein, 3430 EM, the Netherlands. Tel.: 31-(0)88-320-23-00; Fax: 31-(0)88-320-23-99.
E-mail address: orthopedie-research@antoniusziekenhuis.nl (C. Braaksma).

https://doi.org/10.1016/j.joca.2020.08.004
1063-4584/

Widely used PROMs measuring physical functioning in patients with hip or knee complaints are the Hip disability and Osteoarthritis Outcome Score - Physical function Shortform (HOOS-PS)[2] and the Knee Injury and Osteoarthritis Outcome Score - Physical function Shortform (KOOS-PS)[3], respectively. The items on the HOOS-PS and KOOS-PS were selected using Rasch analysis of the Western Ontario McMaster Universities Osteoarthritis Index (WOMAC)[4] and the full-length HOOS[5] and KOOS[6]. The HOOS-PS and KOOS-PS aim to measure physical functioning with fewer items and similar validity compared to the full-length measurements instruments, in order to minimize the burden of the responder and decrease the administrative load. The HOOS-PS and KOOS-PS are selected as outcome measurement instruments by global standard sets of outcome measures, arthroplasty registries and clinical research studies[7–9].

Although the full-length HOOS and KOOS are extensively evaluated, the measurement properties of the short forms of these questionnaires have not been summarized[10–13]. The available systematic reviews did not pool the data quantitatively, included only one article or did not focus on the short form measurement instruments[10–13]. Furthermore, the PROM development and content validity were not qualitatively evaluated. It is important to assess if the HOOS-PS and KOOS-PS are a valid reflection of physical functioning since the outcomes of these measurement instruments are used to evaluate individual patients and to benchmark health care providers.

The goal of this systematic review and meta-analysis is to evaluate all evidence on the measurement properties (content validity, structural validity, internal consistency, reliability, measurement error, cross-cultural validity/measurement invariance, construct validity, criterion validity and responsiveness) and the interpretability of the HOOS-PS and KOOS-PS in patients with hip or knee complaints or undergoing total hip or knee arthroplasty.

Materials and methods

Protocol and registration

This review was reported according to the Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (PRISMA-p)[14]. A study protocol was registered in PROSPERO [CRD42017069539]. The systematic review was conducted according to the.

COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews of PROMs[15]. COSMIN aims to improve the selection of outcome measurement instruments by developing methodology and practical tools for selecting the most suitable outcome measurement instrument.

Eligibility criteria

Eligible studies were full text articles evaluating at least one measurement property or the interpretability of the HOOS-PS and KOOS-PS, or reporting on the development of either the HOOS-PS or KOOS-PS. Furthermore, the development studies of the WOMAC, full-length HOOS or KOOS were eligible, since the items of the HOOS-PS and KOOS-PS were extracted from these measurement instruments in unchanged form. All studies had to evaluate patients of any age with hip or knee complaints or patients who underwent arthroplasty. Included measurement properties were the content validity, structural validity, internal consistency, reliability, measurement error, cross-cultural validity/measurement invariance, construct validity, criterion validity and responsiveness. Table I provides an overview of the definitions of the measurement properties and the interpretability. The HOOS-PS and KOOS-PS had to be patient reported or research administrator assisted. Reviews, study protocols or studies using the outcome measurement instruments for assessment of patients with other limb conditions than hip or knee complaints were excluded. The search was not restricted on language, publication status or study design.

Searches

A literature search was performed in the following electronic bibliographic databases (February 11, 2019): MEDLINE through PubMed, EMBASE through OVID, The Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register), CINAHL and PsychINFO. The search strategy was reviewed by a clinical librarian and can be found in the Supplemental material. References were searched manually to identify other potential studies. Furthermore,

Measurement property	Definition
Content validity	The degree to which the content is an adequate reflection of the construct to be measured
Structural validity	The degree to which the scores are an adequate reflection of the dimensionality of the construct to be measured
Internal consistency	The degree of the interrelatedness among the items
Reliability	The proportion of the total variance in the measurements which is because of true differences among patients
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Cross-cultural validity/measurement invariance	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version
Construct validity	The degree to which the scores are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured
Criterion validity	The degree to which the scores are an adequate reflection of a "gold standard"
Responsiveness	The ability to detect change over time in the construct to be measured
Interpretability	The degree to which one can assign qualitative meaning (that is, clinical or commonly understood connotations) to a PROM's quantitative scores or change in scores.

Table I Taxonomy of the measurement properties and the interpretability, obtained from Mokkink ***et al.*** (2010)[25]

the website http://www.koos.nu was checked for other publications or PhD theses.

Data extraction (selection and coding)

After duplicate removal, two reviewers (CB and NW) identified potentially eligible studies after assessing title and abstract of the retrieved studies independently. If one or both of the reviewers identified a study as potentially eligible, the full text was retrieved and independently assessed by the same two reviewers (CB and NW). Studies were included if they met the eligibility criteria.

The data from the included studies were extracted using a data extract template of the COSMIN manual for systematic reviews of PROMS[16]. The extraction was done by one reviewer (CB) and the second reviewer checked the extracted data (NW) on patient characteristics (number of participants, mean age, sex distribution, disease characteristics, response rate) and type of measurement (HOOS-PS, KOOS-PS, time interval used for follow-up, setting in which the study was conducted, country, language, mode of administration) and all information available on measurement properties. In case differences, consensus was reached by discussion.

Strategy for data synthesis

The methodological quality of the identified studies was assessed per measurement property (taxonomy of measurement properties, Table I) according to the recently updated COSMIN Risk of Bias checklist[17]. Per study, the methodological quality of the measurement property was scored by two independent authors (CB and NW) on a four-point rating scale (i.e., 'very good', 'adequate', 'doubtful' or 'inadequate' quality)[18]. Subsequently, each measurement property was evaluated against the criteria for good measurement properties per study as 'sufficient', 'insufficient' or 'indeterminate'[15]. The quality criteria for good measurement properties are available in the Supplemental materials. A third reviewer was consulted if no consensus was reached (CP).

Summarize quality of evidence and pooling evidence

The overall quality of the PROM was determined using the modified GRADE approach[15], taking into account the methodological quality of the studies and the quality of the measurement properties. The modified GRADE approach was used to downgrade the quality of evidence when there are concerns regarding the risk of bias (evaluated by the COSMIN Risk of Bias checklist), inconsistency in results, imprecision and indirect results[15]. The modified GRADE approach is described in detail in the COSMIN manual for systematic reviews[16]. Quality was graded as 'high', 'moderate', 'low' or 'very low'. The evidence on the measurement properties was pooled quantitatively when the studies were comparable in terms of study population and methodological quality. Otherwise, they were qualitatively summarized. To be able to pool the results of the construct validity and the responsiveness, the authors defined hypotheses about the expected correlations between the HOOS-PS or KOOS-PS and comparator instruments (Table II). All correlations of the (changes in) HOOS-PS and KOOS-PS scores with the comparator instruments found in the included studies were tested against the predefined hypotheses. Afterwards, the percentage of accepted hypotheses and the studies were pooled by calculating the weighted average of the correlations. Discrepancies regarding the pooling of the results and grading of the evidence were resolved by discussion. A third reviewer was consulted when needed (CP).

Statistical analysis

Meta-analysis was done following the method of Feldt and Charter (2006) to compute the pooled internal consistency[19]. Cronbach's alphas were transformed to Fisher's ʒ values that were averaged (weighted average for sample size per study) and converted back to a pooled Cronbach's alpha. Stepwise approach:

1. Calculate a ʒ value per Cronbach's alpha[19]

$$ʒ = 1.1513\ \{\log_{10}\ [(1 + r)/(1 - r)]\}$$

2. Calculate the average weighted ʒ[19].

$$\bar{ʒ} = \Sigma\ (n_j - 3)ʒ_j\ /\ \Sigma\ (n_j - 3)$$

3. Convert the ʒ value back to a pooled Cronbach's alpha[19]

$$r = (10^{\,ʒ/1.1513} - 1)/(10^{\,ʒ/1.513} + 1)$$

Number	Hypothesis
1	Correlations with (changes in) instruments measuring physical function like the physical function subscale of the WOMAC, the KOOS/HOOS and the Oxford Hip Score (OHS)/Oxford Knee Score (OKS) should be >0.50
2	Correlations with (changes in) instruments measuring pain (like the pain subscale of either the WOMAC, OKS/OHS or KOOS/HOOS) or stiffness (like the WOMAC stiffness subscale) should be 0.30–0.50
3	Correlations with (changes in) instruments measuring unrelated constructs like mental health or social functioning should be <0.30
4	Correlations with (changes in) instruments measuring similar constructs should differ by a minimum of 0.10 from correlations with (changes in) instruments measuring related but dissimilar constructs
5	Correlations with (changes in) instruments measuring related constructs should differ by a minimum of 0.10 from correlations with (changes in) instruments measuring unrelated constructs

Table II Predefined hypotheses: the expected correlations between the HOOS-PS or KOOS-PS and comparator instruments

We combined the framework of DerSimonian (1986)[20] and Feldt and Charter (2006)[19] to compute the pooled test-retest reliability. Fisher's transformation to z values were computed by the method of DerSimonian[20]. Computing weighted average was done for the ICC and the confidence interval (95%) the same as for the Cronbach's alpha, with the method of Feldt[19]. Stepwise approach:

1. Calculate the ζ value per ICC[20]

$$\zeta = 0.5 \times \ln((1 + \mathrm{ICC})/(1 - \mathrm{ICC}))$$

2. Calculate the average weighted z[19].

$$\bar{\zeta} = \Sigma\,(n_j - 3)\zeta_j/\Sigma\,(n_j - 3)$$

3. Convert the z value back to a pooled ICC[19]

$$r = (10^{\,z/1.1513} - 1)/(10^{\,z/1.513} + 1)$$

Results

The results of the literature search and selection of the studies are displayed in the PRISMA flow diagram [Fig. 1]. The characteristics of the included PROMs are presented in Table III. The characteristics of the included studies and their populations are presented in Table IV. The summary of findings for each measurement property is presented in Table V.

Content validity

The way PROMs are developed effects the content validity. The HOOS-PS and KOOS-PS were developed via Rasch analysis of the full-length HOOS, KOOS and WOMAC and tested in populations of all ages, from several countries with a wide spectrum of severity of osteoarthritis. The construct to be measured and the target population were clearly described[2,3]. However, no theoretical framework was used to define the construct in a broader setting. The items of the outcome measurement instruments were created in the development studies of the full-length versions and selected in

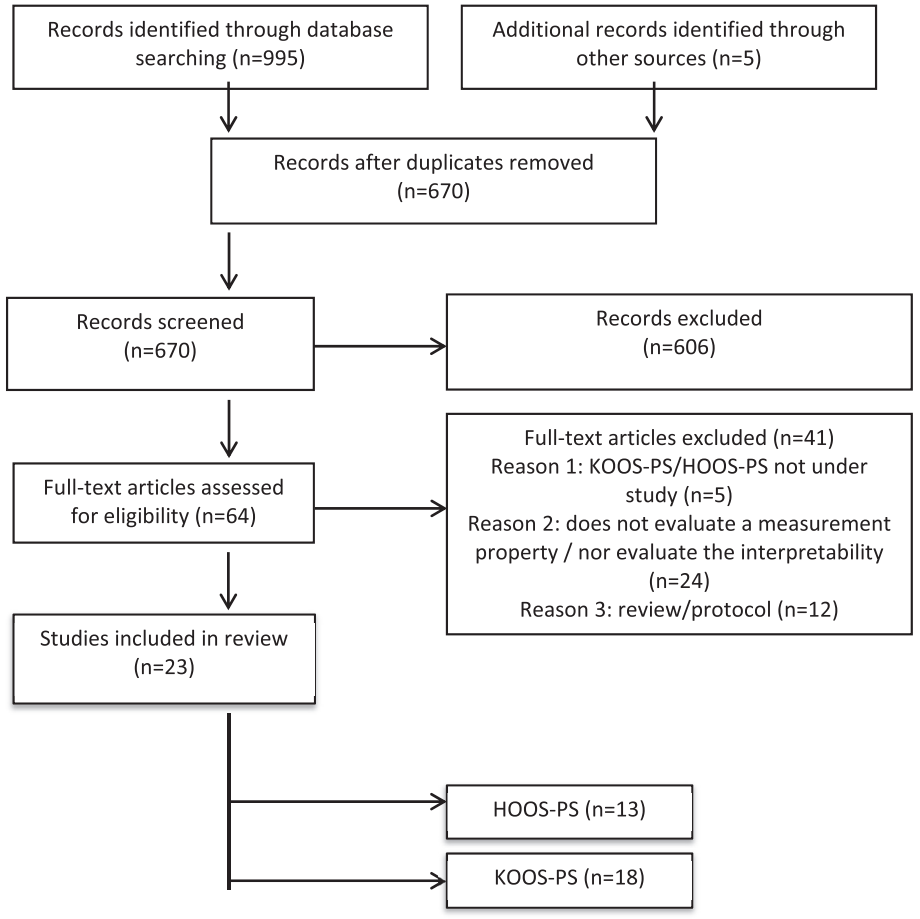

Fig. 1 PRISMA flow diagram of the literature search and selection of the studies.

Measurement property	HOOS-PS(2)	KOOS-PS(3)
Construct	Physical function	Physical function
Target population	People with hip problems	People with knee problems
Mode of administration	Self-administered	Self-administered
Recall period	1 week	1 week
Scale (number of items)	1[5]	1[7]
Response options	None/Mild/Moderate/Severe/Extreme	None/Mild/Moderate/Severe/Extreme
Range of scores/scoring	0-100 (with 0 representing extreme difficulty)	0-100 (with 0 representing extreme difficulty)
Original language	English	English
Available translations	Danish Dutch* English, French* German Italian Norwegian Polish Portuguese (Brazil) Swedish Turkish*	Arabic (Saudi Arabia) Chinese Danish Dutch* English French* German Hindi (India) Italian Korean Norwegian Polish Portuguese* Portuguese (Brazil) Singapore (English) Spanish Swedish Turkish*

* Validated translations.

Table III Characteristics of the included PROMs

Osteoarthritis and Cartilage

unchanged form. Therefore, the methodology and possible limitations of the PROM development studies of the full-length HOOS, KOOS and WOMAC effect the methodological quality of the shorter versions. The items in the development studies were created based on literature review, consulting expert panels and pilot studies[4–6]. No cognitive interviews were conducted to evaluate the comprehensiveness or comprehensibility.

The content validity was inconsistent of the HOOS-PS[2,4,5,21,22] and KOOS-PS[3,4,21,23,24]. Content validity refers to the relevance (the degree to which the content is considered applicable for measuring physical functioning), comprehensiveness (the degree to which all key aspects of the constructs are covered) and the comprehensibility (the degree to which the items, response options and instructions are understood by patients as intended)[25]. Consecutively the relevance, comprehensiveness and comprehensibility are discussed.

There was low quality evidence for an inconsistent relevance of the items of the HOOS-PS and KOOS-PS. The full-length HOOS and KOOS development studies evaluated the items on relevance in patients and included the items with the highest responses. However, they did not use a cut-off value for inclusion of the items[5,6]. The relevance of the items of the HOOS-PS and KOOS-PS was determined in patients undergoing hip or knee arthroplasty and was considered insufficient[21]. In this study including more than 1,200 patients, the item 'running' of the HOOS-PS was found unimportant by 77.7% of patients[21]. In the same study, the items 'kneeling' and 'squatting' of the KOOS-PS were found unimportant by 32.7% and 39.5% of the patients, respectively[21]. The appropriateness of the response options and recall period, and the relevance for the construct of interest and the context of use were not evaluated.

There was very low quality evidence for an insufficient comprehensiveness of the HOOS-PS and KOOS-PS. As no studies evaluated the comprehensiveness, this rating is based on the reviewers rating solely.

There was moderate quality evidence for a sufficient comprehensibility of the HOOS-PS and KOOS-PS. Evidence regarding the comprehensibility is available from studies translating or developing the items of the full-length HOOS, KOOS and WOMAC[4,5,21–24]. The WOMAC development study evaluated the comprehensibility and relevance of a part of the items of the HOOS-PS and KOOS-PS, however not all items[4].The translated full-length HOOS and KOOS into Dutch were rated as comprehensible in a sample of 15 patients per study[22,23], however, methodological quality of these studies was doubtful. It is not clear if skilled group interviewers were used or an appropriate interview guide, if the interviews were recorded and transcribed, how the data was evaluated and analysed and if (besides the items) the instructions and response options were evaluated as well.

Structural validity

There was high quality evidence for a sufficient structural validity of the HOOS-PS. The PROM development study (methodological quality rated as 'very good') assessed the structural validity in a sample of 2,643 persons[2]. Confirmatory factor analysis (CFA) showed a unidimensional construct and showed there was no clustering (location item mean 0 (SD 1.64), X^2 42.29 with a probability of 0.0672, PSI 0.80).

There was moderate quality evidence for an inconsistent structural validity of the KOOS-PS. The KOOS-PS was developed using a Rasch analysis (methodology rated as 'very good')[3] and showed with CFA that the one factor (unidimensional) structure has an adequate fit (location item mean 0 (SD 1.229), X^2 73.34 with a probability of 0.1751, PSI 0.904). Two studies repeated the analysis of the items. First, Franchignoni *et al.* (methodology rated as 'adequate') could not replicate the selection of items of the KOOS-PS in patients with knee osteoarthritis[26]. The items "Twisting/pivoting on your injured knee" showed a borderline infit value and "Rising from bed" showed overfit and thus did not fit the Rasch model. Second, Harris *et al.*, 2013 (methodology rated as 'very good') showed with CFA that there was no acceptable evidence to support the structural validity of the KOOS-PS in 113 knee osteoarthritis patients[27].

Reference	Country; evaluated language	Setting	single-centre or multi-centre study	PROM	Number of patients	Age in years Mean(SD), range	Diagnosis	Gender	FU
Bond 2012[38]	USA; English	Interview-administered	multicentre	HOOS-PS	48	60.3(9.4)	Hip OA, conservative treatment	68.8% female	13 weeks
				KOOS-PS	156	61.2(9.2)	Knee OA, conservative treatment	68.8% female	13 weeks
Davis 2009[28]	Canada; English	Patient-administered, setting unclear	multicentre	HOOS-PS	201	62.3(12.1)	Hip OA, pre and post THR	53% female	6 months
				KOOS-PS	248	64.5(10.3)	Knee OA, pre and post TKR	63% female,	6 months
Davis 2008[2]	Countries: Canada, Sweden, Austria, Finland, France, Germany, Hungary, Iceland, Italy, Poland, Spain, Sweden, Switzerland, United Kingdom; multiple languages.	Patient-administered, setting unclear	multicentre	HOOS-PS	2,991	range 19–96	pre-THR surgery cohorts, community cohort	male: female 1:1.23	NA
Franchignoni 2013[26]	Italy; Italian	Patient-administered, setting unclear	single centre	KOOS-PS	200	69.4(9.5), range 50–84	Knee OA	73.5% female	NA
Goncalves 2010[31]	Portugal; Portuguese	Patient-administered, setting unclear	multicentre	KOOS-PS	85	65.7(6.9)	Knee OA	74.1% female	48 h, 4 weeks or 6 weeks
Gul 2013[32]	Turkey; Turkish	Unclear setting and administration	single centre	KOOS-PS	80	58.9(8.7), range 42–76	Knee OA	88.7% female	NA
Harris 2013[27]	England; English	Patient-administered paper by mail	single centre	KOOS-PS	134	59(11)	Knee OA	50% female	3 months
Mahler 2016[41]	Netherlands; Dutch	Patient-administered paper by mail	single centre	KOOS-PS	161	59(9)	Knee OA	61% female	3 months
Mehta 2016[29]	Sweden, UK, Australia Canada, Czech republic, France, Netherlands; multiple languages	Patient-administered, setting unclear	multicentre	HOOS-PS	745	64.9(11.4)	Hip OA	57% female	NA
				KOOS-PS	1,064	66.8(10.6)	Knee OA	58% female	NA
Ornetti 2009[36]	France; French	Baseline unclear, follow-up patient-administered setting unclear	single centre	HOOS-PS	50	65(10)	Hip OA	74% female	Up to 1 month
				KOOS-PS	87	72(9)	Knee OA	71% female	Up to 1 month
Paulsen 2014[37]	Denmark; Danish	Baseline unclear, follow-up patient-administered paper by mail	multicentre	HOOS-PS	1,335	68, range 23–94	1,175 hip OA, 45 other arthritis, 30 childhood hip diseases, 6 sequel from fracture, 7 necrosis of femoral head	54% female	1 year
Perrucio 2008[3]	Sweden, Canada, France, Estonia, Netherlands; multiple languages	Patient-administered, setting unclear	multicentre	KOOS-PS	2,145	range 26–95	community, knee OA, medial wedge, pre-osteotomy, post ACL	male: female 1:1.4	NA
Ruyssen 2011[30]	France; French	Patient-administered, setting unclear	single centre	HOOS-PS	172 validity 33 reliability 107 responsiveness	65.1(12.3) 64.7(12.1) 65.6(10.2)	Hip OA	53.5% female 63.6% female 48.6% female	12 weeks
				KOOS-PS	128 validity 30 reliability 60 responsiveness	70.9 (10.5) 69.3 (10.9) 71 (10.3)	Knee OA	72.7% female 66.7% female 68.3% female	12 weeks

Singh 2014[35]	USA; English	Patient-administered, setting unclear	multicentre	HOOS-PS	54	Hip and knee cohort together: 60.8 (11.4)	Hip OA	Hip and knee cohort together: 42.6% female	Up to 20 days
				KOOS-PS	141	Hip and knee cohort together: 60.8 (11.4)	Knee OA	Hip and knee cohort together: 42.6% female	Up to 20 days
Stratford 2014[40]	Canada; English	Patient-administered, setting unclear	single centre	KOOS-PS	377	64.4(10.5)	Knee OA	63% female	NA
Wiering 2017[21]	Netherlands; Dutch	Patient-administered online and patient-administered paper	multicentre	HOOS-PS	1,393	72(9.1)	Post THA	Hip and knee cohort together: 65.7% female	NA
				KOOS-PS	1,278	72(9.1)	Post TKA	Hip and knee cohort together: 65.7% female	NA
Yilmaz 2014[34]	Turkey; Turkish	Literate patients: patient-administered, setting unclear. Illiterate patients were read aloud by an investigator.	single centre	HOOS-PS	50	59.1(9.2), range 41–77	Hip OA	74% female	1 week
Groot 2008[23]	Netherlands; Dutch	Patient-administered, setting unclear	unclear	KOOS-PS	15	unclear	Knee OA	unclear	NA
Groot 2007[22]	Netherlands; Dutch	Patient-administered, setting unclear	unclear	HOOS-PS	15	unclear	Hip OA	unclear	NA
Roos 1998[24]	Sweden, USA; Swedish and English	Patient-administered, setting unclear	unclear	KOOS	75	56, range 35–76	Knee OA	not described	NA
Klassbo 2003[5]	Sweden; Swedish	Patient-administered, setting unclear	single centre	HOOS	52	64, range 42–48	Hip complaints, patients with and without hip OA	female/male 35/17	NA
Bellamy 1986[4]	Canada; English	90 face to face interview-administered, 10 telephone interview administered, 15 patients unclear	multicentre	WOMAC	100 (11 hip, 57 knee and 32 both hip and knee)	61.07, range 27–93	Hip or knee OA	female/male 63/37	NA
Gandek 2019[33]	USA; English	Patient administered, either paper-pencil or on internet, at the outpatient clinic or at home	Multicentre	KOOS-PS	1,295	66.5, range 37–100	Knee OA	68.2% female	

Abbreviations: OA: osteoarthritis, TKR: total knee replacement, THR: total hip replacement, NA: not applicable.

Table IV Characteristics of the included study populations

Osteoarthritis and Cartilage

Content validity	Summary (methodologic rating)	Overall rating	Quality of evidence
HOOS-PS[2,4,5,21,22]	Inconsistent relevance (very low), insufficient comprehensiveness (very low) and sufficient comprehensibility (moderate). None of the included studies evaluated all domains of content validity.	Inconsistent	No grading, since overall rating was inconsistent it is not possible to judge quality of evidence
KOOS-PS[3,4,21,23,24]	Inconsistent relevance (very low), insufficient comprehensiveness (very low) and sufficient comprehensibility (moderate). None of the included studies evaluated all domains of content validity.	Inconsistent	No grading, since overall rating was inconsistent it is not possible to judge quality of evidence
Structural validity	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[2]	Unidimensional scale	Sufficient	High as there was one study with very good methodology
KOOS-PS[2,26,27]	Three studies with inconsistent results varying from an unidimensional structure to no acceptable evidence to support the structural validity	Inconsistent	Moderate as there was inconsistency in results
Internal consistency	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[2,28–30]	Pooled Cronbach's alpha = 0.80; total sample size 3761	Sufficient	High as there were several studies with very good methodology
KOOS-PS[26–31,33,34]	Pooled Cronbach's alpha = 0.85; total sample size 3212	Indeterminate	Moderate as the structural validity was inconsistent
Cross-cultural validity	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS	No info available	No info available	No info available
KOOS-PS	No info available	No info available	No info available
Reliability	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[30,34–36]	Pooled ICC = 0.86 (0.67–0.91); total sample size 142	Sufficient	Moderate as there was very serious risk of bias (all studies doubtful methodology)
KOOS-PS[30–32,35,36]	Pooled ICC = 0.81 (0.67–0.87); total sample size 291	Sufficient	Moderate as there was very serious risk of bias (all studies doubtful methodology)
Measurement error	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[36,37]	LoA < MIC	Sufficient	Low as there was very serious risk of bias (only one study with doubtful methodology)
KOOS-PS[27,35,37]	Inconsistent results	Indeterminate	Low as there was serious risk of bias (two studies with doubtful methodology) and there were inconsistent results
Hypotheses testing	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[28–30,34,36,38]	3 out of 5 results in accordance with hypotheses	Insufficient	High: there were several studies with adequate methodology. As the hypotheses came from inadequate comparator instruments, we ignored these results
KOOS-PS[27–32,36,38,40]	3 out of 5 results in accordance with hypotheses	Insufficient	High: there were several studies with adequate methodology. As the hypotheses came from inadequate comparator instruments, we ignored these results
Responsiveness	**Summary or pooled result**	**Overall rating**	**Quality of evidence**
HOOS-PS[27,28,30,31,36,38,41]	No data available of studies with an adequate methodology	Indeterminate	Very low: as there were only studies with inadequate methodology and there were inconsistent results
KOOS-PS[27,41]	2 out of 5 results in accordance with hypotheses	Insufficient	High: as we included two studies with very good methodology and these had consistent results

Abbreviations: ICC: intraclass correlation, LoA: limit of agreement, MIC: minimally important change.

Table V Summary of findings

Internal consistency

There was high quality evidence for a sufficient internal consistency of the HOOS-PS. Pooled Cronbach's alpha in four studies with good methodological quality was 0.80 for the HOOS-PS in 3761 patients[2,28–30].

There was moderate quality evidence for an indeterminate internal consistency of the KOOS-PS. Since we showed that the KOOS-PS is not unidimensional, the internal consistency (pooled outcome of the Cronbach's alpha (0.85 in 3212 patients[26–33]) is difficult to interpret and could not be used and the overall rating was scored as indeterminate. One study was excluded from pooling, because of doubtful methodological quality[34]. One study was rated as sufficient after a discussion within the research team, despite of a Cronbach's alpha of 0.69[30].

Reliability

There was moderate quality evidence for a sufficient reliability of the HOOS-PS and KOOS-PS. Pooled ICC of the HOOS-PS was 0.86 (95% CI 0.67–0.91) in 142 patients[30,34–36]. Pooled ICC of the KOOS-PS was 0.81 (95% CI 0.67–0.87) in 291 patients[30–32,35,36]. Major reasons for the moderate quality evidence were the inclusion of less than fifty subjects per study and not being clear if test conditions and the situation of the patients were similar at baseline and retest. One study was rated as sufficient after consensus meeting, despite an ICC of 0.66[35].

Measurement error

There was low quality evidence of a sufficient measurement error of the HOOS-PS. Limits of agreement (LoA)[36] were smaller than the minimally important change (MIC) obtained from another included study[37] so the measurement error was rated sufficient.

There was low quality evidence for an inconsistent measurement error of the KOOS-PS. The measurement error could not be pooled because there were inconsistent results between studies, probably explained by methodological flaws. The first study showed that the standard error of measurement of 6.7 and an anchor based MIC of 12 results in a smallest detectable change of 18.6 points. This is larger than the MIC, so the measurement error was insufficient[27]. The second study showed that the LoA was smaller than the MIC so the rating was sufficient (no absolute numbers available for the LoA, MIC 28 obtained from another study)[35,36].

Cross-cultural validity|measurement invariance

Cross-cultural validity and measurement invariance could not be evaluated, because no studies evaluated this measurement property of either the HOOS-PS or the KOOS-PS.

Hypotheses testing for construct validity

There was high quality evidence for an insufficient construct validity of the HOOS-PS and KOOS-PS. 60% of the results were in accordance with the hypotheses of both the HOOS-PS and KOOS-PS (Table II); this is below the threshold of 75% for a sufficient rating (Table VI).

Six studies determined the construct validity of the HOOS-PS by correlations with comparator measurement instruments, containing a total of 20 correlations[28–30,34,36,38]. 60% of the results were in accordance with the hypotheses (3 out of 5).

The construct validity of the KOOS-PS was evaluated in nine studies, with a total of 35 correlations of the KOOS-PS with comparator measurement instruments[27,29–32,36,38–40]. 60% of the results were in accordance with the hypotheses (3 out of 5).

	Construct validity		Responsiveness
	HOOS-PS	KOOS-PS	KOOS-PS
1	*Accepted* 100% of the correlations with instruments measuring physical function had a correlation of ≥0.50	*Accepted* 92% of the correlations with instruments measuring physical function had a correlation of ≥0.50	*Rejected* 33% of the correlations with instruments measuring physical function had a correlation of ≥0.50
2	*Rejected* 14% of the correlations with instruments measuring pain, stiffness or a combination of physical function and pain were 0.30–0.50	*Rejected* 8% of the correlations with instruments measuring pain, stiffness or a combination of physical function and pain were 0.30–0.50	*Rejected* 33% of the correlations with instruments measuring pain or a combination of physical function and pain were 0.30–0.50
3	*Rejected* 60% of the correlations with instruments measuring unrelated constructs like mental health or social functioning were <0.30	*rejected* 27% of the correlations with instruments measuring unrelated constructs like mental health or social functioning were <0.30	*Accepted* 100% of the correlations with instruments measuring unrelated constructs like mental health or self-efficacy were <0.30
4	*Accepted* The mean correlation with instruments measuring similar constructs differed 0.156 from the mean correlation with instruments measuring related but dissimilar constructs	*accepted* The mean correlation with instruments measuring similar constructs differed 0.13 from the mean correlation with instruments measuring related but dissimilar constructs	*Rejected* Mean correlation of instruments measuring similar constructs differed 0.06 from the mean correlation of instruments measuring related but dissimilar constructs
5	*Accepted* Mean correlation with instruments measuring related but dissimilar constructs differed 0.34 from instruments measuring unrelated constructs	*accepted* Mean correlation with instruments measuring related but dissimilar constructs differed 0.27 from instruments measuring unrelated constructs.	*Accepted* Mean correlation with instruments measuring related but dissimilar constructs differed 0.32 from instruments measuring unrelated constructs.

Table VI Hypotheses testing for construct validity and responsiveness

	Weighted average score (SD; *n*)			Anchor based values				
	Osteo-arthritis	Post THR/TKR	Post conservative treatment	MIC	PASS	SDC	Floor/ceiling effects	Missing items
HOOS-PS	56.7 (20; 4084)[21,28–30,36–38]	20.1 (19; 2949)[21,28,37])	41.3 (16.2; 20)[36]	23 (CI 19–30)[37]	88 (CI 87–88)[37]	NR	None[34,36]	0–3%[34,36,37]
KOOS-PS	52.9 (17.6; 4651)[27–32,36,38,40,41]	34 (16.6; 2289)[21,28,33]	38.4 (18.4; 257)[27,31,36]	2.2 (SD 17.5) and 12.0[27,35]	NR	16 and 28.3[27,35]	<0.01% ceiling, <2.4% floor; none; ceiling 5%, floor 0.4%;[33,36,40]	0%; 0%; 7–11.7%, squatting and kneeling items missing 4–6% post TKR[31,36]

Abbreviations: SD = standard deviation, *n* = number of patients, THR/TKR = total hip replacement, total knee replacement, NR = not reported, PASS = patient acceptable symptom state, MIC = minimally important change, CI = confidence interval, SDC = smallest detectable change.

Table VII Interpretability: average scores, floor and ceiling effects, MIC/PASS/SDC values for HOOS-PS and KOOS-PS

Osteoarthritis and Cartilage

Feasibility aspects	HOOS-PS	KOOS-PS
Patients comprehensibility	Not evaluated, assumed to be good	Not evaluated, assumed to be good
Clinician's comprehensibility	Good	Good
Type and ease of administration	Self-administered, easy to use	Self-administered, easy to use
Length of the instrument	Short, 5 items	Short, 7 items
Completion time	Not registered, assumed to be maximal 3 min	Not registered, assumed to be maximal 3 min
Patient's required mental and physical ability level	Usage >13 years, mentally competent, all patients with hip complaints	Usage >13 years, mentally competent, all patients with knee complaints
Ease of standardization	No data available	No data available
Ease of score calculation	Easy	Easy
Copyright	Permission not required to use the HOOS-PS	Permission not required to use the KOOS-PS
Cost of an instrument	Free of charge	Free of charge
Required equipment	Paper or online	Paper or online
Availability in different settings	Self-administered. No interview or phone formats are available	Self-administered. No interview or phone formats are available
Regulatory agency's requirement for approval	Not known	Not known

Table VIII Feasibility of the HOOS-PS and KOOS-PS, table based on the COSMIN manual and the guideline for selecting PROMs for Core Outcome Sets[16]

Osteoarthritis and Cartilage

Criterion validity

Criterion validity could not be evaluated, because no studies compared the KOOS-PS or HOOS-PS with summed full-length HOOS or KOOS function and sports subscales.

Responsiveness

There was very low quality evidence for indeterminate responsiveness of the HOOS-PS. All studies used the standardized response mean (SRM) to evaluate the responsiveness of the HOOS-PS[27,28,30,31,36,38,41]. The SRM can be used as an indirect measure when the expected change in health status is known, however it is not the preferred method. Since the expected change in health status on the construct of interest is not known, the SRM cannot be used for evaluating responsiveness of the HOOS-PS[15].

There was high quality evidence for insufficient responsiveness of the KOOS-PS in patients with knee osteoarthritis receiving conservative treatment. Two studies with a very good methodology were pooled[27,41]. 40% of the results were in accordance with the hypotheses (Table VI). Both included studies assessed the correlations between changes of the KOOS-PS with comparator measurement instruments, with predefined hypotheses in patients with knee osteoarthritis receiving conservative treatment. 13 correlations of changes in the KOOS-PS with comparator measurement instruments were found. All other studies evaluated responsiveness had an inadequate methodology and were excluded[27,28,30,31,33,36,38,41] because of using an inappropriate measure of responsiveness.

Interpretability

Table VII presents the summary of the interpretability. It shows the weighted average score and standard deviation on the HOOS-PS and KOOS-PS in patients with osteoarthritis, after total joint replacement or conservative treatment. Furthermore, the MIC, the smallest detectable change and the patient acceptable symptom state are presented. There were no floor or ceiling effects.

Feasibility

Table VIII shows an overview of the feasibility. The authors described the application of the measurement instruments as easy to use, short, and free of charge and copyright.

Discussion

The present study determined the current evidence on the measurement properties of the HOOS-PS and KOOS-PS. The most important finding was the observed lack of several components of the validity of the HOOS-PS and KOOS-PS, such as content validity and construct validity. This implies that the scores on the HOOS-PS and KOOS-PS may inadequately reflect physical functioning in patients with hip or knee complaints. Furthermore, there is evidence for insufficient construct validity and responsiveness in patients with knee osteoarthritis receiving conservative treatment.

All outcome scores and data on measurement properties must be interpreted with caution because the content validity of both outcome measurement instruments was inconclusive. This means that it is unclear if the HOOS-PS and KOOS-PS adequately reflect physical functioning. This can be explained by concerns regarding the relevance and the comprehensiveness of the items of the questionnaires. The unclear content validity can possibly interfere with outcomes on all other measurement properties and should be taken into account when evaluating and interpreting them.

An implication of the problematic validity is the assumption that the HOOS-PS is a reliable outcome measurement instrument, however it cannot be confirmed that the HOOS-PS is reliably measuring the construct physical functioning solely and comprehensively. The found correlations between the HOOS-PS and KOOS-PS with instruments measuring different constructs like pain and stiffness were higher than hypothesized; indicating that they may be measuring a broader construct than just physical functioning. For example, constructs of physical functioning, pain and stiffness may theoretically be distinguishable; however, patients may respond globally. Regarding the HOOS-PS and KOOS-PS, it is possible that the difficulty during activity experienced by patients is influenced by the degree of pain, physical functioning or stiffness.

The inability to distinguish between pain and physical functioning was demonstrated for other outcome measurement instruments[42,43]. Concerning the KOOS-PS, of the eight measurement properties, most were rated as indeterminate or inconsistent and only the reliability was sufficient. This could be due to the inconsistent evidence on content validity.

This is the first review evaluating the responsiveness of the KOOS-PS with studies using adequate methodology, whereas earlier reviews only considered inadequate measures for responsiveness as SRM and effect size[10,13]. The responsiveness was rated as insufficient, indicating that the KOOS-PS is limited in detecting improvement in physical function over time in patients with knee osteoarthritis receiving conservative treatment and thus is probably not the most suitable instrument for measuring outcomes in this population. A previous review evaluating the structural validity concluded that the majority of evidence suggested a unidimensional structure of the KOOS-PS[13], however we were unable to find evidence to support this. Evidence for the structural validity of the KOOS-PS was rated as inconsistent and because of this, we rated the evidence for internal consistency as indeterminate.

A strength of this systematic review is that this review was conducted according to the COSMIN guideline for systematic reviews[15]. The research question was answered extensively and completely. As the HOOS-PS and KOOS-PS are widely used measurement instruments, we could evaluate more evidence on the measurement properties than previous reviews. We evaluated the PROM development and content validity in a systematic and qualitative manner and used not only the PROM development studies of the short forms but also the item development studies of the full-length HOOS, KOOS and WOMAC to obtain information.

The limitations of this study were in particular caused by the inadequate methodology of the included studies. Many measurement properties were evaluated with an inadequate methodology making them unusable to include in this review and more importantly, the inadequate methodology can lead to incorrect conclusions in the studies in question. For instance, with regard to reliability several studies included less than 50 patients. None of the studies used an adequate method to evaluate cross-cultural validity, despite several translations intended to[34], therefore, cross-cultural validity could not be evaluated. Of all included articles, six articles extracted the HOOS-PS and/or KOOS-PS scores out of the full-length HOOS or KOOS. This may have influenced the outcome or missing data. While internal consistency, reliability, construct validity and responsiveness were assessed frequently, the properties content validity, structural validity, cross-cultural validity and criterion validity were not.

Further research should evaluate the content validity of the HOOS-PS and KOOS-PS in more detail and focus on improving the relevance and comprehensiveness of the items to better measure the construct of physical functioning. Alternatively, other PROMs can be explored, for example the promising computer adaptive testing with Patient-Reported Outcomes Measurement Information System (PROMIS)[44]. Although not encouraged, when developing new measurement instruments, it is recommended to use a theoretical model (for example the International Classification of Functioning (ICF[45])) to map the construct to a model. Furthermore, future studies assessing measurement properties of PROMs, are recommended to use the COSMIN design checklist as a guide for achieving adequate methodology[17].

This review and meta-analysis shows that the widespread use in clinical practice of the HOOS-PS and KOOS-PS is not scientifically supported. Although we found evidence for sufficient reliability, the inconsistent evidence on content validity and the insufficient construct validity and responsiveness has implications on all these properties. It may be that the HOOS-PS and KOOS-PS may not measure physical functioning solely and comprehensively. Concluding, scores on the HOOS-PS and KOOS-PS may inadequately reflect physical functioning in patients undergoing total hip and total knee arthroplasty and in patients with hip or knee complaints. Possible consequences of continuing using these questionnaires, are incorrect interpretation of the outcome scores of the individual patients and average outcome scores of healthcare providers with possible patient- and hospital related consequences. Using the HOOS-PS or KOOS-PS as outcome measurement instruments for comparing outcomes, measuring improvements or benchmarking in patients with hip or knee complaints or undergoing arthroplasty should only be done with great caution.

Contributions

Conception and design of the study: CB, NW, SP, RV, RO; Analysis and interpretation of the data: CB, NW, SP, RV, RO; Drafting of the article: CB, NW; Critical revision of the article for important intellectual content: CB, NW, SP, RV, RO, Dr. C.B. Terwee; Final approval of the article: CB, NW, SP, RV, RO; statistical expertise: CB, NW, SP, RO, Prof. Dr. HCW de Vet; Collection and assembly of data: CB, NW, SP.

Christel Braaksma (christelbraaksma@hotmail.com) and Nienke Wolterbeek (orthopedie-research@antoniusziekenhuis.nl) take responsibility for the integrity of the work as a whole, from inception to finished article.

Conflict of interest

None of the authors had conflicts of interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors did not have any other financial interests what could lead to a conflict of interest regarding to this study.

Acknowledgments

The authors want to thank Dr. Associate Prof. C.B. Terwee, associate professor in Measurement, department of epidemiology and biostatistics at Amsterdam University Medical Centre, location VUmc for her critically appraising and improving this manuscript. Furthermore, the authors want to thank Prof. Dr. H.C.W. de Vet, professor in clinimetrics at the VU university medical centre Amsterdam for her feedback, especially checking the methodology of the meta-analysis.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2020.08.004.

References

1. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167.
2. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, *et al*. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthr Cartil 2008 May;16(5):551–9.
3. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, *et al*. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthr Cartil 2008 May;16(5):542–50.

4. Bellamy N, Buchanan WW. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin Rheumatol 1986 Jun;5(2):231–41.
5. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score: an extension of the Western Ontario and McMaster Universities osteoarthritis Index. Scand J Rheumatol 2003;32(1):46–51.
6. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS) - development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998 Aug;28(2):88–96.
7. Dahlberg LE. ICHOM Standard Set for monitoring knee and hip osteoarthritis. Osteoarthr Cartil 2016 Apr;24:S436–7.
8. Aveline C, Roux A Le, Hetet H Le, Gautier JF, Vautier P, Cognet F, *et al*. Pain and recovery after total knee arthroplasty: a 12-month follow-up after a prospective randomized study evaluating nefopam and ketamine for early rehabilitation. Clin J Pain 2014 Sep;30(9):749–54.
9. Gossec L, Paternotte S, Maillefert JF, Combescure C, Conaghan PG, Davis AM, *et al*. The role of pain and functional impairment in the decision to recommend total joint replacement in hip and knee osteoarthritis: an international cross-sectional study of 1909 patients. Report of the OARSI-OMERACT Task Force on total joint replacement. Osteoarthr Cartil 2011 Feb;19(2):147–54.
10. Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, *et al*. A systematic review of measurement properties of patient-reported outcome measures used in patients undergoing total knee arthroplasty. J Arthroplasty 2017 May;32(5):1688–1697.e7.
11. Gagnier JJ, Huang H, Mullins M, Marinac-Dabić D, Ghambaryan A, Eloff B, *et al*. Measurement properties of patient-reported outcome measures used in patients undergoing total hip arthroplasty: a systematic review. JBJS Rev 2018 Jan;6(1):e2.
12. Peer MA, Lane J. The knee injury and osteoarthritis outcome score (KOOS): a review of its psychometric properties in people undergoing total knee arthroplasty. J Orthop Sports Phys Ther 2013 Jan;43(1):20–8.
13. Collins NJ, Prinsen CAC, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. Osteoarthr Cartil 2016 Aug;24(8):1317–29.
14. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: elaboration and explanation. BMJ 2015 Jan 2;350:g7647.
15. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147–57.
16. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW, *et al*. COSMIN Methodology for Systematic Reviews of Patient-Reported Outcome Measures (PROMs) User Manual [Internet] 2018 [cited 2020 Feb 19]. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf.
17. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, *et al*. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1171–9.
18. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the cosmin checklist. Qual Life Res 2012 May;21(4):651–7.
19. Feldt LS, Charter RA. Averaging internal consistency reliability coefficients. Educ Psychol Meas 2006 April;66(2):215–27.
20. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177–88.
21. Wiering B, de Boer D, Delnoij D. Asking what matters: the relevance and use of patient-reported outcome measures that were developed without patient involvement. Health Expect 2017 Dec;20(6):1330–41.
22. de Groot IB, Reijman M, Terwee CB, Bierma-Zeinstra SMA, Favejee M, Roos EM, *et al*. Validation of the Dutch version of the hip disability and osteoarthritis outcome score. Osteoarthr Cartil 2007 Jan;15(1):104–9.
23. de Groot IB, Favejee MM, Reijman M, Verhaar JAN, Terwee CB. The dutch version of the knee injury and osteoarthritis outcome score: a validation study. Health Qual Life Out 2008 Feb 26;6:16.
24. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and osteoarthritis outcome score (KOOS) - validation of a Swedish version. Scand J Med Sci Sports 2007;8(6):439–48.
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737–45.
26. Franchignoni F, Salaffi F, Giordano A, Carotti M, Ciapetti A, Ottonello M, *et al*. Rasch analysis of the 22 knee injury and osteoarthritis outcome score-physical function items in Italian patients with knee osteoarthritis. Arch Phys Med Rehabil 2013 Mar;94(3):480–7.
27. Harris KK, Dawson J, Jones LD, Beard DJ, Price AJ. Extending the use of PROMs in the NHS-using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: a validation study. BMJ Open 2013 Aug;3(8), e003365.
28. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert J-F, *et al*. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthr Cartil 2009 Jul;17(7):843–7.
29. Mehta SP, Sankar A, Venkataramanan V, Lohmander LS, Katz JN, Hawker GA, *et al*. Cross-cultural validation of the ICOAP and physical function short forms of the HOOS and KOOS in a multi-country study of patients with hip and knee osteoarthritis. Osteoarthr Cartil 2016 Dec;24(12):2077–81.
30. Ruyssen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M, *et al*. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. Clin Exp Rheumatol 2011;29(2):231–7.
31. Gonçalves RS, Cabri J, Pinheiro JP, Ferreira PL, Gil J. Reliability, validity and responsiveness of the Portuguese version of the knee injury and osteoarthritis outcome score - physical function short-form (KOOS-PS). Osteoarthr Cartil 2010 Mar;18(3):372–6.
32. Gul ED, Yilmaz O, Bodur H. Reliability and validity of the Turkish version of the knee injury and osteoarthritis outcome score-physical function short-form (KOOS-PS). J Back Musculoskelet Rehabil 2013;26(4):461–6.
33. Gandek B, Roos EM, Franklin PD, Ware JE. Item selection for 12-item short forms of the knee injury and osteoarthritis outcome score (KOOS-12) and hip disability and osteoarthritis outcome score (HOOS-12). Osteoarthr Cartil 2019;27(5):746–53.

34. Yilmaz O, Gul ED, Bodur H. Cross-cultural adaptation and validation of the Turkish version of the hip disability and osteoarthritis outcome score-physical function short-form (HOOS-PS). Rheumatol Int 2014 Jan;34(1):43–9.
35. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol 2014 Mar;41(3):509–15.
36. Ornetti P, Perruccio AV, Roos EM, Lohmander LS, Davis AM, Maillefert JF, *et al.* Psychometric properties of the French translation of the reduced KOOS and HOOS (KOOS-PS and HOOS-PS). Osteoarthr Cartil 2009 Dec;17(12):1604–8.
37. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. Acta Orthop 2014 Feb;85(1):39–48.
38. Bond M, Davis A, Lohmander S, Hawker G. Responsiveness of the OARSI-OMERACT osteoarthritis pain and function measures. Osteoarthr Cartil 2012 Jun;20(6):541–7.
39. Davis AM, Badley EM, Hogg-Johnson S, Ibrahim S, Perruccio AV, Wong R, *et al.* Understanding early recovery following primary total hip and knee replacement. Arthritis Rheum 2009;60:1938.
40. Stratford PW, Kennedy DM. A comparison study of KOOS-PS and KOOS function and sport scores. Phys Ther 2014 Nov;94(11):1614–21.
41. Mahler E, Cuperus N, Bijlsma J, Vliet Vlieland T, van den Hoogen F, den Broeder AA, *et al.* Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis. Scand J Rheumatol 2016 Nov;45(6):518–27.
42. Terwee CB, van der Slikke RMA, van Lummel RC, Benink RJ, Meijers WGH, de Vet HCW. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. J Clin Epidemiol 2006 Jul;59(7):724–31.
43. Stratford P, Kennedy D, Clarke H. Confounding pain and function: the WOMAC's failure to accurately predict lower extremity function. Arthroplast Today 2018 Oct 15;4(4):488–92.
44. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014 May;67(5):516–26.
45. Jette AM. Towards a common language for functioning, disability and health. Phys Ther 2002 May;86(5):726–34.